MATHEMATICS OF COMPUTATION, VOLUME 30, NUMBER 135
JULY 1976, PAGES 641–645

On Asymptotic Properties of Aliquot Sequences

By P. Erdös

Abstract. Put $s^{(1)}(n) = \sigma(n) - n$, $\sigma(n) = \Sigma_{d|n} d$. $s^k(n) = s^{(1)}(s^{(k-1)}(n))$. In this note we prove that for every k the density of integers satisfying

$$s^k(n) = (1 + o(1))n((\sigma(n) - n)/n)^k$$

is 1. Several unsolved problems are stated.

Denote by $\sigma(n)$ the sum of the divisors of n. Define

$$s^0(n) = n, \qquad s^{k+1}(n) = \sigma(s^k(n)) - s^k(n).$$

Catalan and Dickson conjectured that the sequence $s^k(n)$, $k = 1, 2, \ldots$, is always bounded, i.e. either $s^k(n) = 1$ for some k or the sequence becomes periodic. It is a curious fact for which nobody seems to have an explanation that relatively few cycles of size greater than two have been found and none of size three. The Lehmers, and Guy and Selfridge, made extensive numerical investigations. As one consequence of their work, the Catalan-Dickson conjecture is now verified for $n < 276$. Guy and Selfridge have various convincing heuristic arguments which seem to indicate that the Catalan-Dickson conjecture is in fact false. The nicest way of disproving the Catalan-Dickson conjecture would be to find an n so that for every k,

$$s^k(n) > s^{k-1}(n). \tag{1}$$

It seems likely that such an n does not exist, but there does not seem to be much hope of deciding this question.

H. W. Lenstra proved that for every k there is an m so that ($s^1(m)$ will for simplicity be denoted by $s(m)$)

$$s^0(m) < s(m) < \cdots < s^k(m). \tag{2}$$

As far as I know, the proof of Lenstra is unpublished; and since it is very short, I give his proof here:

Let p_i be the ith prime ($p_1 = 2$). It is easy to construct a sequence $(t_i)_{i=1}^{\infty}$ of natural numbers t_i with the property that

(*) $p_i^{t_i+1} | \sigma(p_{i+1}^{t_{i+1}})$ for $i \geqslant 1$ and $t_1 = 2$ (define for instance $t_1 = 2$, $t_{i+1} = \phi(p_i^{t_i+1}(p_{i+1} - 1))$ for $i \geqslant 1$, where ϕ is Euler's ϕ-function).

Received October 10, 1975.
AMS (MOS) subject classifications (1970). Primary 10AXX, 10A20, 10A99.

We define for $l \geqslant 1$: $A_l = \{m \mid m \text{ natural number and } p_i^{t_i} \| m \text{ for } 1 \leqslant i \leqslant l\}$. Here $p^\alpha \| n$ means $p^\alpha \mid n$ and $p^{\alpha+1} \nmid n$. Then

(**) for $m \in A_l$, $t \geqslant 2$, we have $s(m) > m$, and $s(m) \in A_{l-1}$.

Proof of (**). From $l \geqslant 2$ it follows that $12 \mid m$, hence $s(m) > m$. Furthermore, $m \in A_l$, hence $m = p_1^{t_1} \cdots p_t^{t_l} B$ with $(B, m/B) = 1$ and $s(m) = \sigma(m) - m = \sigma(p_1^{t_1}) \cdots \sigma(p_t^{t_l})\sigma(B) - p_1^{t_1} \cdots p_t^{t_l} B$. Now for $1 \leqslant i \leqslant l-1$ we have $p_i^{t_i} \| m$ and (by use of (*)) $p_i^{t_i+1} \mid \sigma(p_{i+1}) \mid \sigma(m)$ which implies $p_i^{t_i} \| s(m)$; conclusion: $s(m) \in A_{l-1}$.

Repeated application of (**) yields $m < s(m) < \cdots < s^{l-1}(m)$ $(\in A_1)$ for $m \in A_l$. Q.E.D.

In the present note we prove the following sharper result:

THEOREM 1. *For every k and $\delta > 0$ and for all n except a sequence of density* 0

$$(3) \qquad (1-\delta)n\left(\frac{s(n)}{n}\right)^i < s^i(n) < (1+\delta)n\left(\frac{s(n)}{n}\right)^i, \qquad 1 \leqslant i \leqslant k.$$

Before we prove our theorem we make a few remarks. First of all, since $s(n)/n \geqslant 7/5$ for all $n \equiv 0(30)$, the lower bound of (3) clearly strengthens (2).

It would be very desirable to strengthen Theorem 1 by showing that (3) remains true if k tends to infinity (not too slowly) together with n, e.g. for $k = (\log n)^\epsilon$. I do not see how this can be done.

Guy and Selfridge have fairly convincing heuristic arguments that for infinitely many values of m, (2) holds for $k < (\log m)^{1-\epsilon}$. I see no way of proving this, but the problem does not seem to be completely hopeless.

The lower bound in Theorem 1 we will prove in full detail; we will only outline the complicated proof of the upper bound.

Before we start our proof we make a few simple remarks which we will need in our proof. Let $S_1, S_2, \ldots, S_k$ be k sets of primes and assume that for each j,

$$(4) \qquad \sum_{p \in S_j} \frac{1}{p} = \infty.$$

Then it easily follows from the sieve of Eratosthenes that almost all integers (i.e. all integers if we neglect a sequence of density 0) have a prime factor $p_y \in S_j$. This result is well known and we leave the simple proof to the reader.

LEMMA 1. *Let t and k be integers. Then almost all integers n have k prime factors $q_1, \ldots, q_k$ satisfying*

$$q_1 \equiv -1 \pmod{t}, \qquad q_j \equiv -1 \pmod{q_{j-1}^2}, \qquad 1 \leqslant j \leqslant k-1.$$

The lemma follows immediately from the previous remark and the classical theorem of Dirichlet. Denote by S_1 the set of primes satisfying $p_1 \equiv -1 \pmod{t}$ and S_j is the set of primes $p_j \equiv -1 \pmod{p_{j-1}^2}$ where $p_{j-1} \in S_j$. The theorem of Dirichlet implies that (4) is satisfied; thus our lemma follows.

Define $f_l(n) = \prod_{p^\alpha \| n; p \leqslant l} p^\alpha$.

LEMMA 2. *For every k and l and almost all n,*

$$(5) \qquad f_l(n) = f_l(s^i(n)) \quad \text{for } i \leqslant k.$$

We shall show that for every $\eta > 0$ and $x > x_0(\eta)$ the number of integers $n < x$ for which (5) does not hold is less than ηx. Choose $u = u(\eta, l)$ so large that $u_0 > l$ and so that for every $x > x_0(\eta)$ there are fewer than $x\eta/2$ integers $n < x$ which are divisible by a prime power $p^\alpha > u$ with $\alpha > 1$. This clearly can be done by choosing u so large that

$$\sum_{p^\alpha > u;\alpha > 1} \frac{1}{p^\alpha} < \frac{\eta}{2}.$$

Now choose $L = u!$; then

$$f_l(n) \mid L \tag{6}$$

for all $n < x$ except the at most $\eta x/2$ integers excluded above.

Put $t = L^2$ in Lemma 1; then $p_i^2 \nmid n$, $1 \leqslant i \leqslant k$, and hence by Lemma 1,

$$s(n) = \sigma(n) - n \equiv 0 \quad \left(\mathrm{mod} \prod_{j=1}^{k-1} p_i\right). \tag{7}$$

But

$$s(n) \not\equiv 0 \pmod{p_j^2}, \qquad 1 \leqslant j \leqslant k-1, \tag{8}$$

since $\sigma(n) \equiv 0 \pmod{L^2 \prod_{j=1}^{k-1} p_j^2}$. Also,

$$f_l(s(n)) = f_l(n), \tag{9}$$

since if $p^\alpha \| n$, $p \leqslant l$ then $p^{2\alpha} \mid \sigma(n)$ by $\sigma(n) \equiv 0 \pmod{L^2}$. (By (6), $p^\alpha \mid t$.)

The same argument gives for every i, $1 \leqslant i \leqslant k$,

$$s^{i+1}(n) = \sigma(s^i(n)) - s^i(n) \equiv 0 \quad \left(\mathrm{mod} \prod_{j=1}^{l-i-1} p_j\right), \tag{7'}$$

$$s^{i+1}(n) \not\equiv 0 \pmod{p_j^2}, \qquad 1 \leqslant j \leqslant k-i-1, \qquad \sigma(s^i(n)) \equiv 0 \pmod{L^2}, \tag{8'}$$

and

$$f_l(s^{i+1}(n)) = f_l(s^i(n)) = f_l(n), \tag{9'}$$

which proves (5) and Lemma 2.

Write

$$\sigma_l(n) = \sum_{d \mid f_l(n)} \frac{n}{d}.$$

LEMMA 3. *For every ϵ and η and $l > l_0(\epsilon, \eta)$ the number of integers $n < x$ for which $\sigma_l(n) > (1-\epsilon)\sigma(n)$ is greater than $(1-\eta)x$.*

We evidently have

$$\sum_{n=1}^{x} (\sigma(n) - \sigma_l(n)) < \sum_{d > l} \sum_{n \leqslant x; n \equiv 0 (\mathrm{mod}\, d)} \frac{n}{d} < \sum_{d > l} \frac{x^2}{d^2} < \frac{x^2}{l}. \tag{10}$$

If there would be ηx integers satisfying $\sigma_l(n) \leqslant (1-\epsilon)\sigma(n)$, we clearly would have

$$\sum_{n=1}^{x} (\sigma(n) - \sigma_l(n)) > \epsilon \sum_{t=1}^{\eta x} t > \epsilon\eta^2 \frac{x^2}{2}. \tag{11}$$

(10) contradicts (11) for $l > 2/\epsilon\eta^2$, which proves Lemma 3.

From Lemmas 2 and 3 we obtain that for all but $\eta x + o(x)$ integers $n < x$ we have that (5) holds and

$$\sigma_l(n) > (1 - \epsilon)\sigma(n). \tag{12}$$

From (5) and (12) we have for every $1 \leqslant i \leqslant k$,

$$\sigma(s^i(n)) \geqslant \sigma_l(s^i(n)) > (1 - \epsilon)s^i(n)\sigma(n)/n, \tag{13}$$

or

$$s^{i+1}(n) = \sigma(s^i(n)) - s^i(n) > (1 - \epsilon)s^i(n)\frac{\sigma(n) - n}{n} = (1 - \epsilon)s^i(n)\frac{s(n)}{n}. \tag{14}$$

From (13) and (14) we immediately obtain that for every $i < k$,

$$s^{i+1}(n) > (1 - \epsilon)^{i+1}\left(\frac{s(n)}{n}\right)^{i+1} n > (1 - \delta)\left(\frac{s(n)}{n}\right)^{i+1} n \quad \text{if } \epsilon < \epsilon(\delta),$$

which completes the proof of the lower bound of (3). It would not be difficult to prove that the lower bound in (3) is valid for $k < \log_r n$ where $\log_r n$ is the r-fold iterated logarithm ($r > 2$), but I do not at present see how to get any reasonable bound for k.

With a little more trouble I can prove that if we neglect a sequence of density 0, then

$$f_l(n) = f_l(s(n)) \tag{15}$$

holds for all $l < \log\log n$, and that this is no longer true for $l = (\log\log n)^{1+\epsilon}$. I do not give the details.

Now I outline the proof of the upper bound of (3). We restrict ourselves to outlining the proof that for almost all integers n,

$$\sigma(s(n))/s(n) < (1 + \epsilon)\sigma(n)/n.$$

The proof is similar for $i > 1$.

In view of Lemma 2 (or (13)) we only have to show that the contribution of the large primes to $\sigma(s(n))/s(n)$ is negligible. This statement easily follows from the

LEMMA 4. *To every $\epsilon > 0$ there is an l so that for all x,*

$$\sum_{n=1}^{x} \sum_{p \mid s(n); p > l} \frac{1}{p} < \epsilon x.$$

Unfortunately, I have at present only a very messy proof of the lemma and this is the reason that I suppress it. I am fairly sure that an elegant and simple proof exists.

Finally I state without proof a few related results. Denote by $f(n)$ the number of $p \mid n$ for which there is another prime $q \mid n$ with $q \equiv 1 \pmod{p}$. Then for almost all integers, $f(n) = (1 + o(1))\log\log\log\log n$. The reason for this weird result is that $p \mid n$ "usually" has the above property if $p < \log\log n$. Similarly, if $F(n)$ denotes the number of $p \mid n$ for which there is a $d \mid n$ satisfying $d \equiv 1 \pmod{p}$, then for almost all n, $F(n) = (1 + o(1))\log\log\log n$.

Denote by $g(n)$ the largest r for which there is a sequence of prime factors p_i of n satisfying $p_{i+1} \equiv 1 \pmod{p_i}$, $1 \leqslant i \leqslant r$, and by $G(n)$ the largest s for which there is a sequence of divisors d_i, $1 \leqslant i \leqslant s$ of n satisfying $d_{i+1} \equiv 1 \pmod{d_i}$, $1 \leqslant i \leqslant s$. Clearly $G(n) \geqslant g(n)$. By the method used in proving Lemma 1 it easily follows that for almost all n, $g(n) \longrightarrow \infty$. On the other hand, $g(n)$ and $G(n)$ tend to infinity very slowly, in fact

$$\frac{1}{x} \sum_{n=1}^{x} G(n) = o(\log_r x)$$

for every r where $\log_r x$ denotes the r times iterated logarithm.

Mathematical Institute
Hungarian Academy of Sciences
Budapest, Hungary

1. E. CATALAN, *Bull. Soc. Math. France*, v. 16, 1887/88, pp. 128–129.

2. L. E. DICKSON, "Theorems and tables on the sum of the divisors of a number," *Quart. J. Math.*, v. 44, 1913, pp. 264–296.

3. RICHARD K. GUY & J. L. SELFRIDGE, "Interim report on aliquot series," *Proc. Manitoba Conf. Numerical Math.* (Univ. Manitoba, Winnipeg, Man., 1971), Dept. Comput. Sci., Univ. Manitoba, Winnipeg, Man., 1971, pp. 557–580. MR 49 #194.

4. RICHARD K. GUY & J. L. SELFRIDGE, "What drives an aliquot sequence?," *Math. Comp.*, v. 29, 1975, pp. 101–107.

5. RICHARD K. GUY, D. H. LEHMER, J. L. SELFRIDGE & M. C. WUNDERLICH, "Second report on aliquot sequences," *Proc. Third Manitoba Conf. Numerical Math.* (Winnipeg, Man., 1973), Utilitas Math., Winnipeg, Man., 1974, pp. 357–368. MR 50 #4455.

6 H. J. J. te RIELE, "A note on the Catalan-Dickson conjecture," *Math. Comp.*, v. 27, 1973, pp. 189–192. MR 48 #3869.